

**COMPOSITIONS AND METHODS OF DIAGNOSING,
MONITORING, STAGING, IMAGING AND TREATING COLON CANCER**

INTRODUCTION

This application claims the benefit of priority from
5 U.S. Provisional Application Serial No. 60/192,667, filed
March 28, 2000.

FIELD OF THE INVENTION

This invention relates, in part, to newly identified
polynucleotides and polypeptides encoded thereby, as well as
10 methods for producing and using these polynucleotides and
polypeptides. Antibodies which are immunospecific for these
polypeptides are also described. Expression of the newly
identified polynucleotides and levels of the polypeptides
encoded thereby are upregulated in or specific to colon cancer
15 tissue. These new polynucleotides and polypeptides, referred
to herein as Colon Cancer Specific Genes or CSGs are believed
to be useful in assays for detecting, diagnosing, monitoring,
staging, prognosticating, imaging and treating cancers,
particularly colon cancer.

20 BACKGROUND OF THE INVENTION

Cancer of the colon is a highly treatable and often
curable disease when localized to the bowel. It is one of the
most frequently diagnosed malignancy in the United States as
well as the second most common cause of cancer death. Surgery
25 is the primary treatment and results in cure in approximately
50% of patients. However, recurrence following surgery is a
major problem and often is the ultimate cause of death.

The prognosis of colon cancer is clearly related to the
degree of penetration of the tumor through the bowel wall and
30 the presence or absence of nodal involvement. These two

characteristics form the basis for all staging systems developed for this disease. Treatment decisions are usually made in reference to the older Duke's or the Modified Astler-Coller (MAC) classification scheme for staging.

5 Bowel obstruction and bowel perforation are indicators of poor prognosis in patients with colon cancer. Elevated pretreatment serum levels of carcinoembryonic antigen (CEA) and of carbohydrate antigen 19-9 (CA 19-9) also have a negative prognostic significance.

10 Age greater than 70 years at presentation is not a contraindication to standard therapies. Acceptable morbidity and mortality, as well as long-term survival, are achieved in this patient population.

15 Because of the frequency of the disease (approximately 160,000 new cases of colon and rectal cancer per year), the identification of high-risk groups, the demonstrated slow growth of primary lesions, the better survival of early-stage lesions, and the relative simplicity and accuracy of screening tests, screening for colon cancer should be a part of routine
20 care for all adults starting at age 50, especially those with first-degree relatives with colorectal cancer.

 Procedures used for detecting, diagnosing, monitoring, staging, and prognosticating colon cancer are of critical importance to the outcome of the patient. For example,
25 patients diagnosed with early colon cancer generally have a much greater five-year survival rate as compared to the survival rate for patients diagnosed with distant metastasized colon cancer. New diagnostic methods which are more sensitive and specific for detecting early colon cancer are clearly
30 needed.

 Colon cancer patients are closely monitored following initial therapy and during adjuvant therapy to determine response to therapy and to detect persistent or recurrent disease of metastasis. There is clearly a need for a colon

cancer marker which is more sensitive and specific in detecting colon cancer, its recurrence, and progression.

Another important step in managing colon cancer is to determine the stage of the patient's disease. Stage
5 determination has potential prognostic value and provides criteria for designing optimal therapy. Generally, pathological staging of colon cancer is preferable over clinical staging because the former gives a more accurate prognosis. However, clinical staging would be preferred were
10 it at least as accurate as pathological staging because it does not depend on an invasive procedure to obtain tissue for pathological evaluation. Staging of colon cancer would be improved by detecting new markers in cells, tissues, or bodily fluids which could differentiate between different stages of
15 invasion.

The present invention relates to newly identified polynucleotides and polypeptides encoded thereby which are referred to herein as Colon Cancer Specific Genes or CSGs, as well as antibodies which are immunospecific for the
20 polypeptides. The present invention also relates to methods for use of these CSGs in detecting, diagnosing, monitoring, staging, prognosticating, imaging and treating colon cancer. For purposes of the present invention, CSG refers, among other things, to native protein expressed by the gene comprising a
25 polynucleotide sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57. By CSG it is also meant herein polynucleotides
30 which, due to degeneracy in genetic coding, comprise variations in nucleotide sequence as compared to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48,
35 49, 50, 51, 52, 53, 54, 55, 56 or 57, but which still encode

the same protein. In the alternative, what is meant by CSG as used herein, means the native mRNA encoded by the gene comprising the polynucleotide sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57, or a polynucleotide which is capable of hybridizing under stringent conditions to the antisense sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57.

Other objects, features, advantages and aspects of the present invention will become apparent to those of skill in the art from the following description. It should be understood, however, that the following description and the specific examples, while indicating preferred embodiments of the invention are given by way of illustration only. Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following description and from reading the other parts of the present disclosure.

SUMMARY OF THE INVENTION

Toward these ends, and others, it is an object of the present invention to provide isolated polynucleotide sequences comprising SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57; fragments or portions of such sequences which contain at least 15 contiguous nucleobases of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38,

39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57; nucleic acid sequences which, due to degeneracy in genetic coding, comprise variations in polynucleotide sequence as compared to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57, but which still encode the same protein; and nucleic acid sequences which are capable of hybridizing under stringent conditions to the antisense sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57.

15 The present invention further relates to isolated polypeptides encoded by the above-described polynucleotides. These polynucleotides and/or the polypeptides encoded thereby are referred to generally herein as Colon Cancer Specific Genes or CSGs.

20 It is another object of the present invention to provide vectors containing the CSG polynucleotides and host cells for expression and recovery of the CSG polypeptides encoded thereby.

It is another object of the present invention to provide 25 antibodies which are immunospecific for CSG polypeptides.

It is another object of the present invention to provide tools for detection of CSGs. Such tools include, but are not limited to, antisense oligonucleotides which specifically hybridize with the CSGs and antibodies which are 30 immunospecific for the CSGs.

It is another object of the present invention to provide a method for diagnosing the presence of colon cancer by analyzing for changes in levels of CSG in cells, tissues or bodily fluids compared with levels of CSG in preferably the 35 same cells, tissues, or bodily fluid type of a normal human

control, wherein a change in levels of CSG in the patient versus the normal human control is associated with colon cancer.

Further provided is a method of diagnosing metastatic
5 cancer in a patient having colon cancer which is not known to have metastasized by identifying a human patient suspected of having colon cancer that has metastasized; analyzing a sample of cells, tissues, or bodily fluid from such patient for CSG; comparing the CSG levels in such cells, tissues, or bodily
10 fluid with levels of CSG in preferably the same cells, tissues, or bodily fluid type of a normal human control, wherein an increase in CSG levels in the patient versus the normal human control is associated with colon cancer which has metastasized.

15 Also provided by the invention is a method of staging colon cancer in a human which has such cancer by identifying a human patient having such cancer; analyzing a sample of cells, tissues, or bodily fluid from such patient for CSG; comparing CSG levels in such cells, tissues, or bodily fluid
20 with levels of CSG in preferably the same cells, tissues, or bodily fluid type of a normal human control sample, wherein an increase in CSG levels in the patient versus the normal human control is associated with a cancer which is progressing and a decrease in the levels of CSG is associated with a
25 cancer which is regressing or in remission.

Further provided is a method of monitoring colon cancer in a human having such cancer for the onset of metastasis. The method comprises identifying a human patient having such cancer that is not known to have metastasized; periodically
30 analyzing a sample of cells, tissues, or bodily fluid from such patient for CSG; comparing the CSG levels in such cells, tissue, or bodily fluid with levels of CSG in preferably the same cells, tissues, or bodily fluid type of a normal human control sample, wherein an increase in CSG levels in the

patient versus the normal human control is associated with a cancer which has metastasized.

Further provided is a method of monitoring the change in stage of colon cancer in a human having such cancer by
5 looking at levels of CSG in a human having such cancer. The method comprises identifying a human patient having such cancer; periodically analyzing a sample of cells, tissues, or bodily fluid from such patient for CSG; comparing the CSG levels in such cells, tissue, or bodily fluid with levels of
10 CSG in preferably the same cells, tissues, or bodily fluid type of a normal human control sample, wherein an increase in CSG levels in the patient versus the normal human control is associated with a cancer which is progressing and a decrease in the levels of CSG is associated with a cancer which is
15 regressing or in remission.

Further provided are new therapeutic agents and methods of identifying therapeutic agents targeted to CSGs for use in imaging and treating disease relating to CSGs such as colon cancer. For example, in one embodiment, therapeutic agents
20 such as antibodies targeted against CSG or fragments of such antibodies can be used to detect or image localization of CSG in a patient for the purpose of detecting or diagnosing a disease or condition. Such antibodies can be polyclonal, monoclonal, or omniclonal or prepared by molecular biology
25 techniques. The term "antibody", as used herein and throughout the instant specification, is also meant to include aptamers and single-stranded oligonucleotides such as those derived from an *in vitro* evolution protocol referred to as SELEX and well known to those skilled in the art. Antibodies
30 can be labeled with a variety of detectable labels including, but not limited to, radioisotopes and paramagnetic metals. Therapeutic agents such as small molecules, antibodies or fragments thereof can also be used in the treatment of diseases characterized by expression of CSG. In these
35 applications, the antibody can be used without or with

derivatization to a cytotoxic agent such as a radioisotope, enzyme, toxin, drug or a prodrug. Therapeutic agents of the present invention also include agonists and antagonists of CSG polypeptides and vaccines capable of inducing an immune
5 response against CSG polypeptides. Such agents can be readily identified in accordance with the teachings herein.

Other objects, features, advantages and aspects of the present invention will become apparent to those of skill in the art from the following description. It should be
10 understood, however, that the following description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only. Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those
15 skilled in the art from reading the following description and from reading the other parts of the present disclosure.

GLOSSARY

The following illustrative explanations are provided to facilitate understanding of certain terms used frequently
20 herein. The explanations are provided as a convenience and are not limitative of the invention.

DIGESTION of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes referred to herein
25 are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan.

For analytical purposes, typically, 1 μg of plasmid or DNA fragment is digested with about 2 units of enzyme in about
30 20 μl of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes.

Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers.

5 Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an
10 agarose or polyacrylamide gel, using well known methods that are routine for those skilled in the art.

GENETIC ELEMENT generally means a polynucleotide comprising a region that encodes a polypeptide or a region that regulates transcription or translation or other processes
15 important to expression of the polypeptide in a host cell, or a polynucleotide comprising both a region that encodes a polypeptide and a region operably linked thereto that regulates expression.

Genetic elements may be comprised within a vector that
20 replicates as an episomal element; that is, as a molecule physically independent of the host cell genome. They may be comprised within mini-chromosomes, such as those that arise during amplification of transfected DNA by methotrexate selection in eukaryotic cells. Genetic elements also may be
25 comprised within a host cell genome; not in their natural state but, rather, following manipulation such as isolation, cloning and introduction into a host cell in the form of purified DNA or in a vector, among others.

ISOLATED means altered "by the hand of man" from its
30 natural state; i.e., that, if it occurs in nature, it has been changed or removed from its original environment, or both.

For example, a naturally occurring polynucleotide or a polypeptide naturally present in a living animal in its natural state is not "isolated," but the same polynucleotide
35 or polypeptide separated from the coexisting materials of its

natural state is "isolated", as the term is employed herein. For example, with respect to polynucleotides, the term isolated means that it is separated from the chromosome and cell in which it naturally occurs.

5 As part of or following isolation, such polynucleotides can be joined to other polynucleotides, such as DNAs, for mutagenesis, to form fusion proteins, and for propagation or expression in a host, for instance. The isolated polynucleotides, alone or joined to other polynucleotides such
10 as vectors, can be introduced into host cells, in culture or in whole organisms. When introduced into host cells in culture or in whole organisms, such DNAs still would be isolated, as the term is used herein, because they would not be in their naturally occurring form or environment. Similarly, the
15 polynucleotides and polypeptides may occur in a composition, such as media formulations, solutions for introduction of polynucleotides or polypeptides, for example, into cells, compositions or solutions for chemical or enzymatic reactions, for instance, which are not naturally occurring compositions,
20 and, therein remain isolated polynucleotides or polypeptides within the meaning of that term as it is employed herein.

LIGATION refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAs. Techniques for ligation are well
25 known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, for instance, Sambrook et al., MOLECULAR CLONING, A LABORATORY MANUAL, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989) and Maniatis et al., pg. 146, as
30 cited below.

OLIGONUCLEOTIDE(S) refers to relatively short polynucleotides of about 8 to about 50 nucleobases. Often the term refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded

ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms.

Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP.

The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

PLASMIDS generally are designated herein by a lower case "p" preceded and/or followed by capital letters and/or numbers, in accordance with standard naming conventions that are familiar to those of skill in the art. Starting plasmids disclosed herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids by routine application of well known, published procedures. Many plasmids and other cloning and expression vectors that can be used in accordance with the present invention are well known and readily available to

those of skill in the art. Moreover, those of skill readily may construct any number of other plasmids suitable for use in the invention. The properties, construction and use of such plasmids, as well as other vectors, in the present invention 5 will be readily apparent to those of skill from the present disclosure.

POLYNUCLEOTIDE(S) generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. Thus, for 10 instance, polynucleotides as used herein refers to, among others, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that 15 may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, polynucleotide as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from 20 different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide.

As used herein, the term polynucleotide includes DNAs or 25 RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotides" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as 30 tritylated bases, to name just two examples, are polynucleotides as the term is used herein.

It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term 35 polynucleotide as it is employed herein embraces such

chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including simple and complex cells, inter alia.

5 POLYPEPTIDES, as used herein, includes all polypeptides as described below. The basic structure of polypeptides is well known and has been described in innumerable textbooks and other publications in the art. In this context, the term is used herein to refer to any peptide or protein comprising two
10 or more amino acids joined to each other in a linear chain by peptide bonds. As used herein, the term refers to both short chains, which also commonly are referred to in the art as peptides, oligopeptides and oligomers, for example, and to longer chains, which generally are referred to in the art as
15 proteins, of which there are many types. It will be appreciated that polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids, and that many amino acids, including the terminal amino acids, may be modified in a given
20 polypeptide, either by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques which are well known to the art. Even the common modifications that occur naturally in polypeptides are too numerous to list exhaustively here, but they are well
25 described in basic texts and in more detailed monographs, as well as in voluminous research literature, and they are well known to those of skill in the art.

Known modifications which may be present in polypeptides of the present invention include, to name an illustrative few,
30 acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking,
35 cyclization, disulfide bond formation, demethylation,

formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, 5 proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well known to those of skill and 10 have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic 15 texts, such as, for instance PROTEINS STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, F., Posttranslational Protein Modifications: Perspectives and 20 Prospects, pgs. 1-12 in POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter et al., Analysis for protein modifications and nonprotein cofactors, Meth. Enzymol. 182: 626-646 (1990) and Rattan et al., Protein Synthesis: 25 Posttranslational Modifications and Aging, Ann. N.Y. Acad. Sci. 663: 48-62 (1992).

It will be appreciated, as is well known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of 30 ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing events and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized

by non-translation natural process and by entirely synthetic methods, as well.

Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and
5 the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance,
10 the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

The modifications that occur in a polypeptide often will be a function of how it is made. For polypeptides made by
15 expressing a cloned gene in a host, for instance, the nature and extent of the modifications in large part will be determined by the host cell posttranslational modification capacity and the modification signals present in the polypeptide amino acid sequence. For instance, as is well
20 known, glycosylation often does not occur in bacterial hosts such as *E. coli*. Accordingly, when glycosylation is desired, a polypeptide should be expressed in a glycosylating host, generally a eukaryotic cell. Insect cells often carry out the same posttranslational glycosylations as mammalian cells and,
25 for this reason, insect cell expression systems have been developed to express efficiently mammalian proteins having native patterns of glycosylation, *inter alia*. Similar considerations apply to other modifications.

It will be appreciated that the same type of modification
30 may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications.

In general, as used herein, the term polypeptide encompasses all such modifications, particularly those that

are present in polypeptides synthesized by expressing a polynucleotide in a host cell.

VARIANT(S) of polynucleotides or polypeptides, as the term is used herein, are polynucleotides or polypeptides that
5 differ from a reference polynucleotide or polypeptide, respectively. Variants in this sense are described below and elsewhere in the present disclosure in greater detail.

A variant may comprise a polynucleotide that differs in nucleotide sequence from another, reference polynucleotide.
10 Generally, differences are limited so that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical.

As noted below, changes in the nucleotide sequence of the variant may be silent. That is, they may not alter the amino
15 acids encoded by the polynucleotide. Where alterations are limited to silent changes of this type a variant will encode a polypeptide with the same amino acid sequence as the reference. Also as noted below, changes in the nucleotide sequence of the variant may alter the amino acid sequence of
20 a polypeptide encoded by the reference polynucleotide. Such nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below.

25 A variant may also comprise a polypeptide that differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference and the variant are closely similar overall and, in many regions, identical.

30 A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.

RECEPTOR MOLECULE, as used herein, refers to molecules
35 which bind or interact specifically with CSG polypeptides of

the present invention and is inclusive not only of classic receptors, which are preferred, but also other molecules that specifically bind to or interact with polypeptides of the invention (which also may be referred to as "binding molecules" and "interaction molecules," respectively and as "CSG binding or interaction molecules". Binding between polypeptides of the invention and such molecules, including receptor or binding or interaction molecules may be exclusive to polypeptides of the invention, which is very highly preferred, or it may be highly specific for polypeptides of the invention, which is highly preferred, or it may be highly specific to a group of proteins that includes polypeptides of the invention, which is preferred, or it may be specific to several groups of proteins at least one of which includes polypeptides of the invention.

Receptors also may be non-naturally occurring, such as antibodies and antibody-derived reagents that bind to polypeptides of the invention.

20 DETAILED DESCRIPTION OF THE INVENTION

Polynucleotides and polypeptides

The present invention relates to newly identified isolated polynucleotides and polypeptides encoded thereby which are upregulated in or specific to colon cancer tissue. These polynucleotides and the polypeptides encoded thereby are believed to be useful as diagnostic markers for cancer, and in particular colon cancer.

For purposes of the present invention, by polynucleotides it is meant to include isolated nucleic acid sequences comprising single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single and double-stranded RNA or hybrids thereof wherein the sequences comprise SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43,

44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57, fragments of at least 15 contiguous nucleobases of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57, nucleic acid sequences which, due to degeneracy in genetic coding, comprise variations in polynucleotide sequence as compared to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57, but which still encode the same protein, and nucleic acid sequences which are capable of hybridizing under stringent conditions to the antisense sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57. By stringent conditions it is meant that hybridization will occur only if there is at least 95%, and more preferably at least 97% identity between the sequences. RNA sequences may be in the form of mRNA while DNA sequences may be in the form of cDNA or genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. As used herein, the term polynucleotide also includes DNAs or RNAs, as described above, that contain one or more modified bases. Examples of modified bases include, but are not limited to, backbone modifications to increase stability and incorporation of unusual bases such as inosine or tritylated bases.

For purposes of the present invention, by polypeptides it is meant to include the recombinant, natural and synthetic polypeptides with amino acid sequences encoded by the polynucleotides of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25,

26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57, or fragments or variants thereof with similar activities and/or levels in cancerous tissues to the amino acid sequences encoded by the polynucleotides of the present invention. Among preferred variants are those that vary from the polypeptides encoded by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57 by conservative amino acid substitutions. Conservative amino acid substitutions typically include replacement, one for another, of the aliphatic amino acids such as Ala, Val, Leu and Ile, the hydroxyl residues Ser and Thr, the acidic residues Asp and Glu, the amide residues Asn and Gln, the basic residues Lys and Arg, and the aromatic residues Phe and Tyr.

Using suppression subtractive hybridization, it has now been found that these polynucleotides, and the polypeptides encoded thereby, are upregulated in, or specific to, colon cancer tissue. Thus, it is believed that these polynucleotides and polypeptides, also referred to herein as Colon Cancer Specific Genes or CSGs, are useful as diagnostic markers for colon cancer, as well as otherwise described herein.

Fragments

Also among preferred embodiments of this aspect of the present invention are polypeptides comprising fragments of CSGs, and fragments of variants and derivatives of the CSGs.

In this regard a fragment is a polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequences encoded by the CSGs of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47,

48, 49, 50, 51, 52, 53, 54, 55, 56 or 57 and variants or derivatives thereof.

Such fragments may be "free-standing," i.e., not part of or fused to other amino acids or polypeptides, or they may be
5 comprised within a larger polypeptide of which they form a part or region. When comprised within a larger polypeptide, the presently discussed fragments most preferably form a single continuous region. However, several fragments may be comprised within a single larger polypeptide. For instance,
10 certain preferred embodiments relate to a fragment of a CSG polypeptide of the present comprised within a precursor polypeptide designed for expression in a host and having heterologous pre- and pro-polypeptide regions fused to the amino terminus of the CSG fragment and an additional region
15 fused to the carboxyl terminus of the fragment. Therefore, fragments in one aspect of the meaning intended herein, refers to the portion or portions of a fusion polypeptide or fusion protein derived from a CSG of the present invention.

As representative examples of polypeptide fragments of
20 the invention, there may be mentioned those which have from about 15 to about 139 amino acids.

In this context "about" includes the particularly recited range and ranges larger or smaller by several, a few, 5, 4, 3, 2 or 1 amino acid at either extreme or at both extremes.
25 Highly preferred in this regard are the recited ranges plus or minus as many as 5 amino acids at either or at both extremes. Particularly highly preferred are the recited ranges plus or minus as many as 3 amino acids at either or at both the recited extremes. Especially preferred are ranges
30 plus or minus 1 amino acid at either or at both extremes or the recited ranges with no additions or deletions. Most highly preferred of all in this regard are fragments from about 15 to about 45 amino acids.

Among especially preferred fragments of the invention are
35 truncation mutants of the CSGs. Truncation mutants include

CSG polypeptides encoded by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, or 57 or variants or derivatives thereof, except for deletion of a continuous series of residues (that is, a continuous region, part or portion) that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus or, as in double truncation mutants, deletion of two continuous series of residues, one including the amino terminus and one including the carboxyl terminus. Fragments having the size ranges set out above also are preferred embodiments of truncation fragments, which are especially preferred among fragments generally.

Also preferred in this aspect of the invention are fragments characterized by structural or functional attributes of the CSGs. Preferred embodiments of the invention in this regard include fragments that comprise alpha-helix and alpha-helix forming regions ("alpha-regions"), beta-sheet and beta-sheet-forming regions ("beta-regions"), turn and turn-forming regions ("turn-regions"), coil and coil-forming regions ("coil-regions"), hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions and high antigenic index regions of CSGs. Garnier-Robson alpha-regions, beta-regions, turn-regions and coil-regions, Chou-Fasman alpha-regions, beta-regions and turn-regions, Kyte-Doolittle hydrophilic regions and hydrophilic regions, Eisenberg alpha and beta amphipathic regions, Karplus-Schulz flexible regions, Emini surface-forming regions and Jameson-Wolf high antigenic index regions are particularly preferred. Among highly preferred fragments in this regard are those that comprise regions of CSGs that combine several structural features, such as several of the features set out above. Such regions may be comprised within a larger polypeptide or may be by themselves a

preferred fragment of the present invention, as discussed above. It will be appreciated that the term "about" as used in this paragraph has the meaning set out above regarding fragments in general.

- 5 Further preferred regions are those that mediate activities of CSGs. Most highly preferred in this regard are fragments that have a chemical, biological or other activity of a CSG, including those with a similar activity or an improved activity, or with a decreased undesirable activity.
- 10 Highly preferred in this regard are fragments that contain regions that are homologs in sequence, or in position, or in both sequence and to active regions of related polypeptides, and which include colon specific-binding proteins. Among particularly preferred fragments in these regards are
- 15 truncation mutants, as discussed above.

It will be appreciated that the invention also relates to, among others, polynucleotides encoding the aforementioned fragments, polynucleotides that hybridize to polynucleotides encoding the fragments, particularly those that hybridize

20 under stringent conditions, and polynucleotides, such as PCR primers, for amplifying polynucleotides that encode the fragments. In these regards, preferred polynucleotides are those that correspondent to the preferred fragments, as discussed above.

25 **Vectors, host cells, expression**

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by

30 recombinant techniques.

Host cells can be genetically engineered to incorporate polynucleotides and express polypeptides of the present invention. For instance, polynucleotides may be introduced into host cells using well known techniques of infection,

35 transduction, transfection, transvection and transformation.

The polynucleotides may be introduced alone or with other polynucleotides. Such other polynucleotides may be introduced independently, co-introduced or introduced joined to the polynucleotides of the invention.

5 Thus, for instance, polynucleotides of the invention may be transfected into host cells with another, separate, polynucleotide encoding a selectable marker, using standard techniques for co-transfection and selection in, for instance, mammalian cells. In this case the polynucleotides generally
10 will be stably incorporated into the host cell genome.

Alternatively, the polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. The vector construct may be introduced into host cells by the aforementioned techniques. Generally, a plasmid vector
15 is introduced as DNA in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. Electroporation also may be used to introduce polynucleotides into a host. If the vector is a virus, it may be packaged in vitro or introduced into a packaging cell and the packaged
20 virus may be transduced into cells. A wide variety of techniques suitable for making polynucleotides and for introducing polynucleotides into cells in accordance with this aspect of the invention are well known and routine to those of skill in the art. Such techniques are reviewed at length
25 in Sambrook et al. cited above, which is illustrative of the many laboratory manuals that detail these techniques. In accordance with this aspect of the invention the vector may be, for example, a plasmid vector, a single- or double-stranded phage vector, or a single- or double-stranded RNA or
30 DNA viral vector. Such vectors may be introduced into cells as polynucleotides, preferably DNA, by well known techniques for introducing DNA and RNA into cells. The vectors, in the case of phage and viral vectors also may be and preferably are introduced into cells as packaged or encapsidated virus by
35 well known techniques for infection and transduction. Viral

vectors may be replication competent or replication defective. In the latter case viral propagation generally will occur only in complementing host cells.

Preferred among vectors, in certain respects, are those
5 for expression of polynucleotides and polypeptides of the present invention. Generally, such vectors comprise cis-acting control regions effective for expression in a host operatively linked to the polynucleotide to be expressed. Appropriate trans-acting factors either are supplied by the host, supplied
10 by a complementing vector or supplied by the vector itself upon introduction into the host.

In certain preferred embodiments in this regard, the vectors provide for specific expression. Such specific expression may be inducible expression or expression only in
15 certain types of cells or both inducible and cell-specific. Particularly preferred among inducible vectors are vectors that can be induced for expression by environmental factors that are easy to manipulate, such as temperature and nutrient additives. A variety of vectors suitable to this aspect of the
20 invention, including constitutive and inducible expression vectors for use in prokaryotic and eukaryotic hosts, are well known and employed routinely by those of skill in the art.

The engineered host cells can be cultured in conventional nutrient media, which may be modified as appropriate for,
25 inter alia, activating promoters, selecting transformants or amplifying genes. Culture conditions, such as temperature, pH and the like, previously used with the host cell selected for expression generally will be suitable for expression of polypeptides of the present invention as will be apparent to
30 those of skill in the art.

A great variety of expression vectors can be used to express a polypeptide of the invention. Such vectors include chromosomal, episomal and virus-derived vectors e.g., vectors derived from bacterial plasmids, from bacteriophage, from
35 yeast episomes, from yeast chromosomal elements, from viruses

such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids, all may be used for expression in accordance with this aspect of the present invention. Generally, any vector suitable to maintain, propagate or express polynucleotides to express a polypeptide in a host may be used for expression in this regard.

10 The appropriate DNA sequence may be inserted into the vector by any of a variety of well-known and routine techniques. In general, a DNA sequence for expression is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction
15 endonucleases and then joining the restriction fragments together using T4 DNA ligase. Procedures for restriction and ligation that can be used to this end are well known and routine to those of skill. Suitable procedures in this regard, and for constructing expression vectors using alternative
20 techniques, which also are well known and routine to those of skill, are set forth in great detail in Sambrook et al. cited elsewhere herein.

The DNA sequence in the expression vector is operatively linked to appropriate expression control sequence(s),
25 including, for instance, a promoter to direct mRNA transcription. Representatives of such promoters include the phage lambda PL promoter, the *E. coli* lac, trp and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name just a few of the well-known
30 promoters. It will be understood that numerous promoters not mentioned are suitable for use in this aspect of the invention are well known and readily may be employed by those of skill in the manner illustrated by the discussion and the examples herein.

In general, expression constructs will contain sites for transcription initiation and termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the mature transcripts expressed by the
5 constructs will include a translation initiating AUG at the beginning and a termination codon appropriately positioned at the end of the polypeptide to be translated.

In addition, the constructs may contain control regions that regulate as well as engender expression. Generally, in
10 accordance with many commonly practiced procedures, such regions will operate by controlling transcription, such as repressor binding sites and enhancers, among others.

Vectors for propagation and expression generally will include selectable markers. Such markers also may be suitable
15 for amplification or the vectors may contain additional markers for this purpose. In this regard, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells. Preferred markers include dihydrofolate reductase
20 or neomycin resistance for eukaryotic cell culture, and tetracycline or ampicillin resistance genes for culturing *E. coli* and other bacteria.

The vector containing the appropriate DNA sequence as described elsewhere herein, as well as an appropriate
25 promoter, and other appropriate control sequences, may be introduced into an appropriate host using a variety of well known techniques suitable to expression therein of a desired polypeptide. Representative examples of appropriate hosts include bacterial cells, such as *E. coli*, *Streptomyces* and
30 *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS and Bowes melanoma cells; and plant cells. Hosts for a great variety of expression constructs are well known, and those of skill will be enabled
35 by the present disclosure readily to select a host for

expressing a polypeptides in accordance with this aspect of the present invention.

More particularly, the present invention also includes recombinant constructs, such as expression constructs, 5 comprising one or more of the sequences described above. The constructs comprise a vector, such as a plasmid or viral vector, into which such a sequence of the invention has been inserted. The sequence may be inserted in a forward or reverse orientation. In certain preferred embodiments in this 10 regard, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and there are many commercially available vectors suitable for use in the present 15 invention.

The following vectors, which are commercially available, are provided by way of example. Among vectors preferred for use in bacteria are pQE70, pQE60 and pQE-9, available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, 20 pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia. Among preferred eukaryotic vectors are PWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. These 25 vectors are listed solely by way of illustration of the many commercially available and well known vectors that are available to those of skill in the art for use in accordance with this aspect of the present invention. It will be appreciated that any other plasmid or vector suitable for, for 30 example, introduction, maintenance, propagation or expression of a polynucleotide or polypeptide of the invention in a host may be used in this aspect of the invention.

Promoter regions can be selected from any desired gene using vectors that contain a reporter transcription unit 35 lacking a promoter region, such as a chloramphenicol acetyl

transferase ("cat") transcription unit, downstream of restriction site or sites for introducing a candidate promoter fragment; i.e., a fragment that may contain a promoter. As is well known, introduction into the vector of a promoter-containing fragment at the restriction site upstream of the cat gene engenders production of CAT activity, which can be detected by standard CAT assays. Vectors suitable to this end are well known and readily available. Two such vectors are pKK232-8 and pCM7. Thus, promoters for expression of polynucleotides of the present invention include not only well known and readily available promoters, but also promoters that readily may be obtained by the foregoing technique, using a reporter gene.

Among known bacterial promoters suitable for expression of polynucleotides and polypeptides in accordance with the present invention are the *E. coli* lacI and lacZ and promoters, the T3 and T7 promoters, the gpt promoter, the lambda PR, PL promoters and the trp promoter. Among known eukaryotic promoters suitable in this regard are the CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous sarcoma virus ("RSV"), and metallothionein promoters, such as the mouse metallothionein-I promoter.

Selection of appropriate vectors and promoters for expression in a host cell is a well known procedure and the requisite techniques for expression vector construction, introduction of the vector into the host and expression in the host are routine skills in the art.

The present invention also relates to host cells containing the above-described constructs discussed above. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection or other methods.

5 Such methods are described in many standard laboratory manuals, such as Davis et al. BASIC METHODS IN MOLECULAR BIOLOGY, (1986).

Constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant
10 sequence. Alternatively, the polypeptides of the invention can be synthetically produced by conventional peptide synthesizers.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of
15 appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook et al., MOLECULAR
20 CLONING: A LABORATORY MANUAL, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

Generally, recombinant expression vectors will include origins of replication, a promoter derived from a highly-expressed gene to direct transcription of a downstream
25 structural sequence, and a selectable marker to permit isolation of vector containing cells after exposure to the vector. Among suitable promoters are those derived from the genes that encode glycolytic enzymes such as 3-phosphoglycerate kinase ("PGK"), a-factor, acid phosphatase,
30 and heat shock proteins, among others. Selectable markers include the ampicillin resistance gene of *E. coli* and the *trp1* gene of *S. cerevisiae*.

Transcription of the DNA encoding the polypeptides of the present invention by higher eukaryotes may be increased by
35 inserting an enhancer sequence into the vector. Enhancers are

cis-acting elements of DNA, usually about from 10 to 300 bp that act to increase transcriptional activity of a promoter in a given host cell-type. Examples of enhancers include the SV40 enhancer, which is located on the late side of the replication origin at bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Polynucleotides of the invention, encoding the heterologous structural sequence of a polypeptide of the invention generally will be inserted into the vector using standard techniques so that it is operably linked to the promoter for expression. The polynucleotide will be positioned so that the transcription start site is located appropriately 5' to a ribosome binding site. The ribosome binding site will be 5' to the AUG that initiates translation of the polypeptide to be expressed. Generally, there will be no other open reading frames that begin with an initiation codon, usually AUG, and lie between the ribosome binding site and the initiating AUG. Also, generally, there will be a translation stop codon at the end of the polypeptide and there will be a polyadenylation signal and a transcription termination signal appropriately disposed at the 3' end of the transcribed region.

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide. The signals may be endogenous to the polypeptide or they may be heterologous signals.

The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals but also additional heterologous functional regions. Thus, for instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and

persistence in the host cell, during purification or during subsequent handling and storage. A region also may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide.

- 5 The addition of peptide moieties to polypeptides to engender secretion or excretion, to improve stability and to facilitate purification, among others, are familiar and routine techniques in the art.

Suitable prokaryotic hosts for propagation, maintenance
10 or expression of polynucleotides and polypeptides in accordance with the invention include *Escherichia coli*, *Bacillus subtilis* and *Salmonella typhimurium*. Various species of *Pseudomonas*, *Streptomyces*, and *Staphylococcus* are suitable hosts in this regard. Moreover, many other hosts also known
15 to those of skill may be employed in this regard.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements
20 of the well known cloning vector pBR322. Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, Wis., USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be
25 expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, where the selected promoter is inducible it is induced by appropriate means (e.g., temperature shift or exposure to
30 chemical inducer) and cells are cultured for an additional period.

Cells typically then are harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification. Microbial
35 cells employed in expression of proteins can be disrupted by

any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well know to those skilled in the art.

5 Various mammalian cell culture systems can be employed for expression, as well. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblast, described in Gluzman et al., Cell 23: 175 (1981). Other cell lines capable of expressing a compatible vector include for
10 example, the C127, 3T3, CHO, HeLa, human kidney 293 and BHK cell lines.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation sites,
15 splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking non-transcribed sequences that are necessary for expression. In certain preferred embodiments in this regard DNA sequences derived from the SV40 splice sites, and the SV40 polyadenylation sites are used for required non-
20 transcribed genetic elements of these types.

The CSG polypeptide can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose
25 chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification. Well known techniques for refolding protein may be employed
30 to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Polypeptides of the present invention include naturally purified products, products of chemical synthetic procedures, and products produced by recombinant techniques from a
35 prokaryotic or eukaryotic host, including, for example,

bacterial, yeast, higher plant, insect and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, 5 polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes.

CSG polynucleotides and polypeptides may be used in accordance with the present invention for a variety of 10 applications, particularly those that make use of the chemical and biological properties of the CSGs. Additional applications relate to diagnosis and to treatment of disorders of cells, tissues and organisms. These aspects of the invention are illustrated further by the following discussion.

15 **Antibodies**

The CSG polypeptides of the invention or their fragments or variants thereof, or cells expressing them can be used as immunogens to produce antibodies immunospecific for the CSG polypeptides. The term "immunospecific" means that the 20 antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art. These antibodies can be polyclonal or monoclonal. In addition, by the term "antibody", it is meant to include chimeric, single chain and 25 humanized and fully human antibodies as well as Fab fragments or products of Fab expression libraries.

Antibodies generated against the CSG polypeptides can be obtained by administering the polypeptides or epitope-bearing fragments, variants or cells to an animal, preferably a 30 nonhuman, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., *Nature* (1975) 256:495-497), the trioma 35 technique, the human B-cell hybridoma technique (Kozbor et

al., *Immunology Today* (1983) 4:72) and the EBV-hybridoma technique (Cole et al., *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, pp. 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies
5 (U.S. Patent No. 4,946,778) can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms including other mammals, can be used to express humanized antibodies.

The above-described antibodies can be used to isolate or
10 to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.

Antibodies against CSG polypeptides can also be used to treat colon cancer, among others.

Diagnostic Tools

15 The present invention also relates to diagnostic tools such as antibodies which are immunospecific for CSGs or labeled oligonucleotide probes which hybridize to CSGs.

Antibodies immunospecific for CSGs are described in detail in the preceding section.

20 Antisense oligonucleotides which hybridize to a portion of a polynucleotide of the present invention can be chemically synthesized via an automated oligonucleotide synthesizer or produced via alternative methods such as *in vitro* recombinant DNA-mediated techniques and by expression of DNA in cells and
25 organisms. Most often oligonucleotides comprise single-stranded deoxyribonucleotides. However, oligonucleotides may also comprise single-or double-stranded ribonucleotide, RNA:DNA hybrids and double-stranded DNAs.

Methods of Use

30 The present invention also relates to assays and methods, both quantitative and qualitative, for detecting, diagnosing, monitoring, staging and prognosticating cancers by comparing levels of CSG in a human patient with those of CSG in a normal human control. For purposes of the present invention, what
35 is meant by "CSG levels" is, among other things, native

protein expressed by a polynucleotide sequence comprising SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57. By "CSG" it is also meant herein polynucleotides which, due to degeneracy in genetic coding, comprise variations in nucleotide sequence as compared to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57, but which still encode the same protein. The native protein being detected may be whole, a breakdown product, a complex of molecules or chemically modified. In the alternative, what is meant by "CSG" as used herein, means the native mRNA encoded by the gene comprising the polynucleotide sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57, or a polynucleotide which is capable of hybridizing under stringent conditions to the antisense sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57. Such levels are preferably determined in at least one of cells, tissues and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing overexpression of CSG protein compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of colon cancer.

All the methods of the present invention may optionally include determining the levels of other cancer markers as well

as CSG. Other cancer markers, in addition to CSG, useful in the present invention will depend on the cancer being tested and are known to those of skill in the art.

Diagnostic Assays

5 The present invention provides methods for diagnosing the presence of cancer, and in particular colon cancer, by analyzing for changes in levels of CSG in cells, tissues or bodily fluids compared with levels of CSG in cells, tissues or bodily fluids of preferably the same type from a normal
10 human control, wherein an increase in levels of CSG in the patient versus the normal human control is associated with the presence of colon cancer.

Without limiting the instant invention, typically, for a quantitative diagnostic assay a positive result indicating
15 the patient being tested has cancer is one in which cells, tissues or bodily fluid levels of the cancer marker, such as CSG, are at least two times higher, and most preferably are at least five times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control.

20 The present invention also provides a method of diagnosing metastatic colon cancer in a patient having colon cancer which has not yet metastasized for the onset of metastasis. In the method of the present invention, a human cancer patient suspected of having colon cancer which may have
25 metastasized (but which was not previously known to have metastasized) is identified. This is accomplished by a variety of means known to those of skill in the art.

In the present invention, determining the presence of CSG levels in cells, tissues or bodily fluid, is particularly
30 useful for discriminating between colon cancer which has not metastasized and colon cancer which has metastasized. Existing techniques have difficulty discriminating between colon cancer which has metastasized and colon cancer which has not metastasized and proper treatment selection is often
35 dependent upon such knowledge.

In the present invention, the cancer marker levels are measured in such cells, tissues or bodily fluid is CSG, and are then compared with levels of CSG in preferably the same cells, tissue or bodily fluid type of a normal human control.

5 That is, if the cancer marker being observed is CSG in serum, this level is preferably compared with the level of CSG in serum of a normal human control. An increase in the CSG in the patient versus the normal human control is associated with colon cancer which has metastasized.

10 Without limiting the instant invention, typically, for a quantitative diagnostic assay a positive result indicating the cancer in the patient being tested or monitored has metastasized is one in which cells, tissues or bodily fluid levels of the cancer marker, such as CSG, are at least two
15 times higher, and most preferably are at least five times higher, than in preferably the same cells, tissues or bodily fluid of a normal patient.

Normal human control as used herein includes a human patient without cancer and/or non cancerous samples from the
20 patient. In the methods for diagnosing or monitoring for metastasis, normal human control may preferably also include samples from a human patient that is determined by reliable methods to have colon cancer which has not metastasized.

Staging

25 The invention also provides a method of staging colon cancer in a human patient. The method comprises identifying a human patient having such cancer and analyzing cells, tissues or bodily fluid from such human patient for CSG. The CSG levels determined in the patient are then compared with
30 levels of CSG in preferably the same cells, tissues or bodily fluid type of a normal human control, wherein an increase in CSG levels in the human patient versus the normal human control is associated with a cancer which is progressing and a decrease in the levels of CSG (but still increased over true

normal levels) is associated with a cancer which is regressing or in remission.

Monitoring

Further provided is a method of monitoring colon cancer
5 in a human patient having such cancer for the onset of metastasis. The method comprises identifying a human patient having such cancer that is not known to have metastasized; periodically analyzing cells, tissues or bodily fluid from such human patient for CSG; and comparing the CSG levels
10 determined in the human patient with levels of CSG in preferably the same cells, tissues or bodily fluid type of a normal human control, wherein an increase in CSG levels in the human patient versus the normal human control is associated with a cancer which has metastasized. In this method, normal
15 human control samples may also include prior patient samples.

Further provided by this invention is a method of monitoring the change in stage of colon cancer in a human patient having such cancer. The method comprises identifying a human patient having such cancer; periodically analyzing
20 cells, tissues or bodily fluid from such human patient for CSG; and comparing the CSG levels determined in the human patient with levels of CSG in preferably the same cells, tissues or bodily fluid type of a normal human control, wherein an increase in CSG levels in the human patient versus
25 the normal human control is associated with a cancer which is progressing in stage and a decrease in the levels of CSG is associated with a cancer which is regressing in stage or in remission. In this method, normal human control samples may also include prior patient samples.

30 Monitoring a patient for onset of metastasis is periodic and preferably done on a quarterly basis. However, this may be done more or less frequently depending on the cancer, the particular patient, and the stage of the cancer.

Prognostic Testing and Clinical Trial Monitoring

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased levels of CSG. The present invention provides a method in which a test sample is obtained from a human patient and CSG is detected. The presence of higher CSG levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly colon cancer.

The effectiveness of therapeutic agents to decrease expression or activity of the CSGs of the invention can also be monitored by analyzing levels of expression of the CSGs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient, or cells as the case may be, to the agent being tested.

Detection of Genetic Lesions or Mutations

The methods of the present invention can also be used to detect genetic lesions or mutations in CSG, thereby determining if a human with the genetic lesion is at risk for colon cancer or has colon cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion and/or addition and/or substitution of one or more nucleotides from the CSGs of this invention, a chromosomal rearrangement of CSG, aberrant modification of CSG (such as of the methylation pattern of the genomic DNA), the presence of a non-wild type splicing pattern of a mRNA transcript of CSG, allelic loss of CSG, and/or inappropriate post-translational modification of CSG protein. Methods to detect such lesions in the CSGs of this invention are known to those of skill in the art.

Assay Techniques

Assay techniques that can be used to determine levels of gene expression (including protein levels), such as CSG of the present invention, in a sample derived from a patient are well known to those of skill in the art. Such assay methods include, without limitation, radioimmunoassays, reverse transcriptase PCR (RT-PCR) assays, immunohistochemistry assays, *in situ* hybridization assays, competitive-binding assays, Western Blot analyses, ELISA assays and proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel based approaches such as mass spectrometry or protein interaction profiling. Among these, ELISAs are frequently preferred to diagnose a gene's expressed protein in biological fluids.

An ELISA assay initially comprises preparing an antibody, if not readily available from a commercial source, specific to CSG, preferably a monoclonal antibody. In addition a reporter antibody generally is prepared which binds specifically to CSG. The reporter antibody is attached to a detectable reagent such as radioactive, fluorescent or enzymatic reagent, for example horseradish peroxidase enzyme or alkaline phosphatase.

To carry out the ELISA, antibody specific to CSG is incubated on a solid support, e.g. a polystyrene dish, that binds the antibody. Any free protein binding sites on the dish are then covered by incubating with a non-specific protein such as bovine serum albumin. Next, the sample to be analyzed is incubated in the dish, during which time CSG binds to the specific antibody attached to the polystyrene dish. Unbound sample is washed out with buffer. A reporter antibody specifically directed to CSG and linked to a detectable reagent such as horseradish peroxidase is placed in the dish resulting in binding of the reporter antibody to any monoclonal antibody bound to CSG. Unattached reporter antibody is then washed out. Reagents for peroxidase

activity, including a colorimetric substrate are then added to the dish. Immobilized peroxidase, linked to CSG antibodies, produces a colored reaction product. The amount of color developed in a given time period is proportional to the amount of CSG protein present in the sample. Quantitative results typically are obtained by reference to a standard curve.

A competition assay can also be employed wherein antibodies specific to CSG are attached to a solid support and labeled CSG and a sample derived from the host are passed over the solid support. The amount of label detected which is attached to the solid support can be correlated to a quantity of CSG in the sample.

Using all or a portion of a nucleic acid sequence of a CSG of the present invention as a hybridization probe, nucleic acid methods can also be used to detect CSG mRNA as a marker for colon cancer. Polymerase chain reaction (PCR) and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction. RT-PCR can thus reveal by amplification the presence of a single species of mRNA. Accordingly, if the mRNA is highly specific for the cell that produces it, RT-PCR can be used to identify the presence of a specific type of cell.

Hybridization to clones or oligonucleotides arrayed on a solid support (i.e. gridding) can be used to both detect the expression of and quantitate the level of expression of that gene. In this approach, a cDNA encoding the CSG gene is fixed

to a substrate. The substrate may be of any suitable type including but not limited to glass, nitrocellulose, nylon or plastic. At least a portion of the DNA encoding the CSG gene is attached to the substrate and then incubated with the
5 analyte, which may be RNA or a complementary DNA (cDNA) copy of the RNA, isolated from the tissue of interest. Hybridization between the substrate bound DNA and the analyte can be detected and quantitated by several means including but not limited to radioactive labeling or fluorescence labeling
10 of the analyte or a secondary molecule designed to detect the hybrid. Quantitation of the level of gene expression can be done by comparison of the intensity of the signal from the analyte compared with that determined from known standards. The standards can be obtained by *in vitro* transcription of the
15 target gene, quantitating the yield, and then using that material to generate a standard curve.

Of the proteomic approaches, 2D electrophoresis is a technique well known to those in the art. Isolation of individual proteins from a sample such as serum is
20 accomplished using sequential separation of proteins by different characteristics usually on polyacrylamide gels. First, proteins are separated by size using an electric current. The current acts uniformly on all proteins, so smaller proteins move farther on the gel than larger proteins.
25 The second dimension applies a current perpendicular to the first and separates proteins not on the basis of size but on the specific electric charge carried by each protein. Since no two proteins with different sequences are identical on the basis of both size and charge, the result of a 2D separation
30 is a square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue
5 biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood.

10 *In Vivo Targeting of CSG/Colon Cancer Therapy*

Identification of these CSGs is also useful in the rational design of new therapeutics for imaging and treating cancers, and in particular colon cancer. For example, in one embodiment, antibodies which specifically bind to a CSG can
15 be raised and used *in vivo* in patients suspected of suffering from colon cancer. Antibodies which specifically bind CSG can be injected into a patient suspected of having colon cancer for diagnostic and/or therapeutic purposes. Thus, another aspect of the present invention provides for a method for
20 preventing the onset and treatment of colon cancer in a human patient in need of such treatment by administering to the patient an effective amount of antibody. By "effective amount" it is meant the amount or concentration of antibody needed to bind to the target antigens expressed on the tumor
25 to cause tumor shrinkage for surgical removal, or disappearance of the tumor. The binding of the antibody to the overexpressed CSG is believed to cause the death of the cancer cell expressing such CSG. The preparation and use of antibodies for *in vivo* diagnosis and treatment is well known
30 in the art. For example, antibody-chelators labeled with Indium-111 have been described for use in the radioimmunoscentigraphic imaging of carcinoembryonic antigen expressing tumors (Sumerdon et al. Nucl. Med. Biol. 1990 17:247-254). In particular, these antibody-chelators have
35 been used in detecting tumors in patients suspected of having

recurrent colorectal cancer (Griffin et al. J. Clin. Onc. 1991 9:631-640). Antibodies with paramagnetic ions as labels for use in magnetic resonance imaging have also been described (Lauffer, R.B. Magnetic Resonance in Medicine 1991 22:339-5 342). Antibodies directed against CSG can be used in a similar manner. Labeled antibodies which specifically bind CSG can be injected into patients suspected of having colon cancer for the purpose of diagnosing or staging of the disease status of the patient. The label used will be selected in 10 accordance with the imaging modality to be used. For example, radioactive labels such as Indium-111, Technetium-99m or Iodine-131 can be used for planar scans or single photon emission computed tomography (SPECT). Positron emitting labels such as Fluorine-19 can be used in positron emission 15 tomography. Paramagnetic ions such as Gadlinium (III) or Manganese (II) can be used in magnetic resonance imaging (MRI). Presence of the label, as compared to imaging of normal tissue, permits determination of the spread of the cancer. The amount of label within an organ or tissue also 20 allows determination of the presence or absence of cancer in that organ or tissue.

Antibodies which can be used in *in vivo* methods include polyclonal, monoclonal and omniclonal antibodies and antibodies prepared via molecular biology techniques. 25 Antibody fragments and aptamers and single-stranded oligonucleotides such as those derived from an *in vitro* evolution protocol referred to as SELEX and well known to those skilled in the art can also be used.

Vaccines

30 Another aspect of the invention relates to compositions and methods for inducing an immunological response in a mammal. In one embodiment, a mammal is inoculated with a CSG polypeptide, or a fragment or variant thereof, in an amount adequate to produce an antibody and/or T cell immune response 35 against CSG polypeptide. In another embodiment, a vector

directing expression of CSG polynucleotide *in vivo* is used to induce such an immunological response and to produce antibody. The immune response against the CSG polypeptide is expected to protect the mammal from diseases, in particular colon
5 cancer.

Thus, the present invention also relates to an immunological/vaccine formulation (composition) which, when introduced into a mammal, induces an immunological response in that mammal to CSG polypeptide wherein the composition
10 comprises a CSG polypeptide, fragment or variant thereof or a vector expressing a CSG gene or fragment thereof. The vaccine formulation may further comprise a suitable carrier. Since CSG polypeptide may be broken down in the stomach, the vaccine formulation is preferably administered parenterally
15 via subcutaneous, intramuscular, intravenous, or intradermal injection. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation
20 isotonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, as sealed ampules and vials, and may be stored in a freeze-dried condition requiring only the addition of the sterile
25 liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will
30 depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Screening Assays

The CSG polypeptides of the present invention can also be employed in screening processes for compounds which
35 activate (agonists) or inhibit activation of (antagonists, or

otherwise called inhibitors) the CSG polypeptides of the present invention. Thus, polypeptides of the invention can be used to identify agonists or antagonists from, for example, cells, cell-free preparations, chemical libraries, and natural
5 product mixtures. These agonists or antagonists may be natural or modified substrates, ligands, receptors, enzymes, etc., as the case may be, of the polypeptides of the present invention; or may be structural or functional mimetics of the polypeptide of the present invention. See Coligan et al.,
10 *Current Protocols in Immunology* 1(2):Chapter 5 (1991).

CSG polypeptides are responsible for various biological functions, including pathologies such as colon cancer. Accordingly, it is desirable to identify compounds which stimulate CSG polypeptides on the one hand (agonists) and
15 which can inhibit the function of CSG polypeptides (antagonists) on the other hand. Agonists and antagonists can be employed for therapeutic and prophylactic purposes for conditions such as colon cancer.

In general, such screening procedures involve using
20 appropriate cells which express the CSG polypeptide or respond to CSG polypeptide of the present invention. Such cells include those from mammals, yeast, *Drosophila* and *E. coli*. Cells which express the CSG polypeptide (or cell membranes containing the expressed polypeptide) or respond to CSG
25 polypeptides are then contacted with a candidate compound to observe binding, or stimulation or inhibition of a functional response. The CSG activity of the cells which were contacted with the candidate compounds is compared with the CSG activity in the same type of cells which were not contacted with the
30 candidate compounds.

The assays may simply test binding of a candidate compound wherein adherence to the cells bearing the CSG polypeptide is detected by means of a label directly or indirectly associated with the candidate compound or in an
35 assay involving competition with a labeled competitor.

Further, these assays can test whether the candidate compound results in a signal generated by activation of the CSG polypeptide using detection systems appropriate to the cells bearing the CSG polypeptide. Inhibitors of activation are
5 generally assayed in the presence of a known agonist and the effect of the candidate compound upon activation by the agonist is observed.

Alternatively, the assays may comprise the steps of mixing a candidate compound with a solution containing a CSG
10 polypeptide to form a mixture, measuring CSG activity in the mixture, and comparing the CSG activity of the mixture to a standard.

The CSG polynucleotide, polypeptides and antibodies immunospecific for the polypeptides can also be used to
15 configure assays for detecting the effect of added compounds on the production of CSG mRNA and polypeptides in cells. For example, an ELISA for measuring secreted or cell associated levels of CSG polypeptide using monoclonal and polyclonal antibodies can be constructed by standard methods known in the
20 art. The ELISA can then be used to discover agents which may inhibit or enhance the production of CSG from suitably manipulated cells or tissues. Standard methods for conducting these screening assays are well understood in the art.

The CSG polypeptides can also be used to identify
25 membrane bound or soluble receptors, if any, through standard receptor binding techniques known in the art. These include, but are not limited to, ligand binding and crosslinking assays in which the CSG is labeled with a radioactive isotope (e.g. ¹²⁵I), chemically modified (e.g. biotinylated), or fused to a
30 peptide sequence suitable for detection or purification, and incubated with a source of the putative receptor (cells, cell membranes, cell supernatants, tissue extracts, bodily fluids). Other methods include biophysical techniques such as surface plasmon resonance and spectroscopy. In addition to being used
35 for purification and cloning of the receptor, these binding

assays can be used to identify agonists and antagonists of CSG which compete with the binding of CSG to receptors. Standard methods for conducting these screening assays are well understood in the art.

5 Examples of potential CSG polypeptide antagonists include, but are not limited to: antibodies; oligonucleotides or proteins which are closely related to the CSGs; ligands, substrates, receptors, and enzymes of the CSG polypeptides; fragment of these ligands, substrates, receptors and enzymes;
10 and small molecules which bind to the polypeptide of the present invention so that the activity of the polypeptide is prevented.

Thus, the present invention also relates to screening kits for identifying agonists, antagonists, ligands,
15 receptors, substrates, enzymes, etc. for CSG polypeptides; or compounds which decrease or enhance the production of CSG polypeptides. Such kits preferably comprise a CSG polypeptide; a recombinant cell expressing a CSG polypeptide or a cell membrane expressing a CSG polypeptide; and an
20 antibody to a CSG polypeptide.

Prophylactic and Therapeutic Methods

This invention also relates to methods of treating abnormal conditions such as, colon cancer, related to both an excess of and insufficient amounts of CSG polypeptide
25 activity.

If the activity of CSG polypeptide is in excess, several approaches are available. One approach comprises administering to a subject an inhibitor compound (antagonist) as hereinabove described along with a pharmaceutically
30 acceptable carrier in an amount effective to inhibit the function of the CSG polypeptide, such as, for example, by blocking the binding of ligands, substrates, enzymes, receptors, etc., or by inhibiting a second signal, and thereby alleviating the abnormal condition. In another
35 approach, soluble forms of CSG polypeptides still capable of

binding the ligand, substrate, enzymes, receptors, etc. in competition with endogenous CSG polypeptide can be administered. Typical embodiments of such competitors comprise fragments of the CSG polypeptide.

- 5 In still another approach, expression of the gene encoding endogenous CSG polypeptide can be inhibited using expression blocking techniques. Known blocking techniques involve the use of antisense sequences, either internally generated or separately administered. See, for example,
- 10 O'Connor, *J Neurochem* (1991) 56:560 in Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988). Alternatively, oligonucleotides which form triple helices with the gene can be supplied. See, for example, Lee et al., *Nucleic Acids Res* (1979) 6:3073; Cooney
- 15 et al., *Science* (1988) 241:456; Dervan et al., *Science* (1991) 251:1360. These oligomers can be administered *per se* or the relevant oligomers can be expressed *in vivo*.

Several approaches are also available for treating abnormal conditions related to an under-expression of CSG and

20 its activity. One approach comprises administering to a subject a therapeutically effective amount of a compound which activates CSG polypeptide, i.e., an agonist as described above, in combination with a pharmaceutically acceptable carrier, to thereby alleviate the abnormal condition.

25 Alternatively, gene therapy can be employed to effect the endogenous production of CSG by the relevant cells in the subject. For example, a polynucleotide of the invention can be engineered for expression in a viral vector such as a replication defective retroviral vector. The retroviral

30 expression construct can then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer

35 cells can be administered to a subject for engineering cells

in vivo and expression of the polypeptide *in vivo*. For an overview of gene therapy, see Chapter 20, *Gene Therapy and other Molecular Genetic-based Therapeutic Approaches*, (and references cited therein) in *Human Molecular Genetics*, T 5 Strachan and A P Read, BIOS Scientific Publishers Ltd (1996). Another approach is to administer a therapeutic amount of CSG polypeptides in combination with a suitable pharmaceutical carrier.

Formulation and Administration

10 Peptides, such as a soluble form of CSG polypeptide, and agonist and antagonist peptides or small molecules, can be formulated in various combinations with suitable pharmaceutical carriers. These formulations comprise a therapeutically effective amount of the peptide or small 15 molecule, and a pharmaceutically acceptable carrier or excipient. Such carriers include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The formulation is selected in accordance with the mode of administration, and is well within 20 the skill of the art. The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

The compositions of present invention can be employed 25 alone or in conjunction with other compounds, such as other therapeutic compounds.

Preferred forms of systemic administration of these pharmaceutical compositions include injection, typically by intravenous injection. Other injection routes, such as 30 subcutaneous, intramuscular, or intraperitoneal, can also be used. Alternative means for systemic administration include transmucosal and transdermal administration using penetrants such as bile salts or fusidic acids or other detergents. In addition, if properly formulated in enteric or encapsulated 35 formulations, oral administration may also be possible. These

compositions can also be administered topically in the form of salves, pastes, gels and the like.

The dosage range required depends on the composition, the route of administration, the nature of the formulation, the nature of the subject's condition, and the judgment of the attending practitioner. Suitable dosages, however, are generally in the range of 0.1-100 μ g of peptide or small molecule per kg of subject. Wide variations in the needed dosage, however, are to be expected in view of the variety of compounds available and the differing efficiencies of various routes of administration. For example, oral administration is expected to require higher dosages than administration by intravenous injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art.

Polypeptides used in treatment can also be generated endogenously in the subject, in treatment modalities often referred to as "gene therapy" as described above. Thus, for example, cells from a subject may be engineered with a CSG polynucleotide of the present invention, such as a DNA or RNA, to encode a polypeptide *ex vivo*, and for example, by the use of a retroviral plasmid vector to encode a polypeptide *in vivo*. The cells are then introduced into the subject.

EXAMPLES

The present invention is further described by the following examples. The examples are provided solely to illustrate the invention by reference to specific embodiments. These exemplifications, while illustrating certain aspects of the invention, does not portray the limitations or circumscribe the scope of the disclosed invention.

The examples are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. Routine molecular biology techniques of the following examples can be carried

out as described in standard laboratory manuals, such as Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

5 **Example 1: Suppression subtractive hybridization**
(Clontech PCR-SELECT)

Clontech PCR-SELECT is a PCR based subtractive hybridization method designed to selectively enrich for cDNAs corresponding to mRNAs differentially expressed between two
10 mRNA populations (Diatchenko et al. *Proc. Natl. Acad. Sci. USA*, Vol. 93, pp. 6025-6030, 1996).

In this method, cDNA is prepared from the two mRNA populations which are to be compared (Tester: cDNA population in which the differentially expressed messages are sought and
15 Driver: cDNA population in which the differentially expressed transcripts are absent or low). The tester sample is separated in two parts and different PCR adapters are ligated to the 5' ends. Each tester is separately annealed to excess driver (first annealing) and then pooled and again annealed
20 (second annealing) to excess driver. During the first annealing, sequences common to both populations anneal. Additionally, the concentration of high and low abundance messages are normalized since annealing is faster for abundant molecules due to the second order kinetics of hybridization.
25 During the second annealing, cDNAs unique or overabundant to the tester can anneal together. Such molecules have different adapters at their ends. The addition of additional driver during the second annealing enhances the enrichment of the desired differentially expressed sequences. During subsequent
30 PCR, molecules that have different adapters at each end amplify exponentially. Molecules which have identical adapters, or adapters at only one end, or no adapters (driver sequences) either do not amplify or undergo linear amplification. The end result is enrichment for cDNAs

corresponding to differentially expressed messages (unique to the tester or upregulated in the tester).

This technique was used to identify transcripts unique to colon tissues or messages overexpressed in colon cancer.

5 Pairs of matched samples isolated from the same patient, a cancer sample, and the "normal" adjacent tissue from the same tissue type were utilized. The mRNA from the cancer tissue is used as the "tester", and the non-cancer mRNA as a "driver". The non-cancer "driver" is from the same individual

10 and tissue as the cancer sample (Matched). Alternatively the "driver" can be from a different individual but the same tissue as the tumor sample (unmatched). In some cases, mixtures of mRNAs derived from non-cancer tissue types different from the cancer tissue type were used as the

15 "driver". This approach allows the identification of transcripts whose expression is specific or upregulated in the cancer tissue type analyzed.

Several subtracted libraries were generated for colon. The product of the subtraction experiments was used to

20 generate cDNA libraries. These cDNA libraries contain Expressed Sequence Tags (ESTs) from genes that are colon cancer specific, or upregulated in colon. Selected clones from each cDNA PCR Select library were sequenced and are depicted as SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,

25 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56 or 57.

Example 2: Semi-quantitative Polymerase Chain Reaction (SQ-PCR)

30 SQ-PCR is a method that utilizes end point PCR on serial dilutions of cDNA samples in order to determine relative expression patterns of genes of interest in multiple samples. Using random hexamer primed Reverse Transcription (RT) cDNA panels are created from total RNA samples. Gene specific

primers are then used to amplify fragments using Polymerase Chain Reaction (PCR) technology from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A
5 positive reaction in the most dilute sample indicates the highest relative expression value. This is determined by analysis of the sample reactions on a 2-4% agarose gel. The tissue samples used include 12 normal, 12 cancer and 6 pairs tissue specific cancer and matching samples.

10 **SEQ ID NO: 6; Clone ID clnc00000953 (Sqcln015)**

Semi quantitative PCR was done using the following primers:

Sqcln015forward:

5' AGGGACCTGAGGCTTTCTTTA 3' (SEQ ID NO:58)

Sqcln015reverse:

15 5' CACCACACCTTGGATTCAAAG 3' (SEQ ID NO:59)

Table 1 shows the absolute numbers which are relative levels of expression of sqcln015 in 12 normal samples from 12 different tissues. These RNA samples are individual samples or are commercially available pools, originated by pooling
20 samples of a particular tissue from different individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive
25 reaction in the most dilute sample indicates the highest relative expression value.

Table 1: Sqcln015 in Normal Tissues

	Tissue	Normal
30	Breast	0
	Colon	1
	Endometrium	0
	Kidney	1
	Liver	0
	Lung	1000
35	Ovary	0
	Prostate	100
	Small Intestine	1

Stomach	0
Testis	10
Uterus	0

Relative levels of expression in Table 1 show that normal lung, prostate and testes exhibit expression of sqcln015. However, colon normal tissue has a very low level of expression.

Table 2 shows the absolute numbers which are relative levels of expression of sqcln015 in 12 cancer samples from 12 different tissues. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 2: Sqcln015 in Cancer Tissue Samples

Tissue	Cancer
Bladder	0
Breast	0
Colon	0
Kidney	1
Liver	0
Lung	1
Ovary	1
Pancreas	0
Prostate	0
Stomach	0
Testes	1
Uterus	1

Relative levels of expression in Table 2 show that sqcln015 is expressed in very low levels in kidney, lung, ovary, testes and uterine carcinomas.

Table 3 shows the absolute numbers which are relative levels of expression of sqcln015 in 6 colon cancer matching samples. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. Using Polymerase Chain Reaction (PCR) technology expression

levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 3: sqcln015 in Colon Cancer Samples

Sample ID	Tissue	Cancer	NAT
9609B019	Colon	100	1
9709C074RA	Colon	0	0
9705F002D	Colon	0	0
9608B012	Colon	0	0
4004709A1	Colon	0	0
9707C004GB	Colon	100	0

Relative levels of expression in Table 3 shows that sqcln015 is expressed in moderate levels in two of the six colon cancer samples with little or no expression in the matching normal adjacent tissue (NAT).

SEQ ID NO:8; Clone ID clnc00000827 (Sqcln017)

Semi quantitative PCR was done using the following primers:

Sqcln017forward:

5' TCTCCTCCCGACAAGAAGTATCT 3' (SEQ ID NO:60)

Sqcln017reverse:

5' GGACCAGGAGCAGCAAATC 3' (SEQ ID NO:61)

Table 4 shows the absolute numbers which are relative levels of expression of sqcln017 in 12 normal samples from 12 different tissues. These RNA samples are individual samples or are commercially available pools, originated by pooling samples of a particular tissue from different individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 4: Sqcln017 in Normal Tissue Samples

	Tissue	Normal
	Breast	1
	Colon	1
5	Endometrium	0
	Kidney	0
	Liver	1
	Lung	1
	Ovary	0
10	Prostate	1
	Small	1
	Stomach	10
	Testis	1
	Uterus	1

15 Relative levels of expression in Table 4 show that normal breast, colon, liver, lung, prostate, small intestine, uterus, stomach and testes exhibit low levels of expression of sqcln015.

Table 5 shows the absolute numbers which are relative
20 levels of expression of sqcln017 in 12 cancer samples from 12 different tissues. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene
25 expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 5: Sqcln017 in Cancer Tissue Samples

	Tissue	Cancer
	Bladder	0
30	Breast	1
	Colon	0
	Kidney	0
	Liver	0
	Lung	0
35	Ovary	0
	Pancreas	0
	Prostate	0
	Stomach	1
	Testes	1
40	Uterus	1

Relative levels of expression in Table 5 show that sqcln017 is expressed in very low levels in breast, stomach, testicular and uterine carcinomas.

Table 6 shows the absolute numbers which are relative levels of expression of sqcln017 in 6 colon cancer matching samples. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 6: Colon Cancer Samples

Sample ID	Tissue	Cancer	NAT
9609B019	Colon	1	1
9709C074RA	Colon	0	0
9705F002D	Colon	1	0
9608B012	Colon	0	0
4004709A1	Colon	100	0
9707C004GB	Colon	0	100

Relative levels of expression in Table 6 show that sqcln017 is expressed in low to moderate levels in three of the six colon cancer samples with low expression in one matching normal adjacent tissue (NAT) and moderate expression in another NAT.

SEQ ID NO: 13; Clone ID clnc00000720 (Sqcln022)

Semi quantitative PCR was done using the following primers:

Sqcln015forward:

5' CTCCCATCGCTCCAGAGTG 3' (SEQ ID NO:62)

Sqcln015reverse:

5' GACCGAAACCAACAGTGATTG 3' (SEQ ID NO:63)

Table 7 shows the absolute numbers which are relative levels of expression of sqcln022 in 12 normal samples from 12 different tissues. These RNA samples are individual samples or are commercially available pools, originated by pooling 5 samples of a particular tissue from different individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive 10 reaction in the most dilute sample indicates the highest relative expression value.

Table 7: Sqcln022 in Normal Tissue Samples

	Tissue	Normal
15	Breast	10
	Colon	10
	Endometrium	10
	Kidney	1
	Liver	1
	Lung	100
20	Ovary	0
	Prostate	0
	Small Intestine	0
	Stomach	1
	Testis	10
25	Uterus	10

Relative levels of expression in Table 7 show that normal lung, breast, endometrium, testis and uterus exhibit moderate to low levels of expression of sqcln022. However, colon normal tissue does not exhibit any expression.

30 Table 8 shows the absolute numbers which are relative levels of expression of sqcln022 in 12 cancer samples from 12 different tissues. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression 35 levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 8: Sqcln022 in Cancer Tissue Samples

	Tissue	Cancer
5	Bladder	0
	Breast	0
	Colon	1
	Kidney	0
	Liver	0
10	Lung	1
	Ovary	1
	Pancreas	1
	Prostate	1
	Stomach	1
	Testes	1
	Uterus	1

15 Relative levels of expression in Table 8 show that sqcln022 is expressed in very low levels in colon, lung, ovary, pancreas, prostate, stomach, testes and uterine carcinomas.

Table 9 shows the absolute numbers which are relative levels of expression of sqcln022 in 6 colon cancer matching
 20 samples. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in
 25 duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 9: Sqcln022 in Colon Cancer Samples

	Sample ID	Tissue	Cancer	NAT
30	9609B019	Colon	0	1000
	9709C074RA	Colon	100	1000
	9705F002D	Colon	100	1000
35	9608B012	Colon	100	100
	4004709A1	Colon	100	10
	9707C004GB	Colon	100	1000

Relative levels of expression in Table 9 show that sqcln022 is expressed in moderate levels in five of the six colon

cancer samples. However good expression is also noticed for the matching normal adjacent tissue (NAT).

SEQ ID NO:14; Clone ID clnc00000666 (Sqcln023)

Semi quantitative PCR was done using the following
5 primers:

Sqcln023forward:

5' GTGGGGGAAAGATGACTAAAATA 3' (SEQ ID NO:64)

Sqcln023reverse:

5' TCCCTTTGCCGTTACCACT 3' (SEQ ID NO:65)

10 Table 10 shows the absolute numbers which are relative levels of expression of sqcln023 in 12 normal samples from 12 different tissues. These RNA samples are individual samples or are commercially available pools, originated by pooling samples of a particular tissue from different individuals.
15 Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest
20 relative expression value.

Table 10: Sqcln023 in Normal Tissue Samples

Tissue	Normal
Breast	1
Colon	1
25 Endometrium	1
Kidney	1
Liver	0
Lung	1
Ovary	0
30 Prostate	10
Small Intestine	100
Stomach	10
Testis	1
Uterus	1

35 Relative levels of expression in Table 10 show that normal small intestine, stomach, and prostate show moderate expression of sqcln023. Low levels of expression are apparent

in normal breast, colon, endometrium, kidney, lung, testis and uterus.

Table 11 shows the absolute numbers which are relative levels of expression of sqcln023 in 12 cancer samples from 12 different tissues. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 11: Sqcln023 in Cancer Tissue Samples

Tissue	Cancer
Bladder	0
Breast	0
Colon	1
Kidney	1
Liver	1
Lung	1
Ovary	1
Pancreas	1
Prostate	1
Stomach	1
Testes	10
Uterus	1

Relative levels of expression in Table 11 show that sqcln023 is expressed in low levels in colon, kidney, liver, lung, ovary, pancreas, prostate, stomach and uterine carcinomas.

Table 12 shows the absolute numbers which are relative levels of expression of sqcln023 in 6 colon cancer matching samples. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 12: sqcln023 in Colon Cancer Samples

Sample ID	Tissue	Cancer	NAT
9609B019	Colon	100	10
9709C074RA	Colon	100	100
9705F002D	Colon	100	100
9608B012	Colon	0	100
4004709A1	Colon	100	1000
9707C004GB	Colon	100	100

Relative levels of expression in Table 12 show that sqcln023 is expressed in moderate levels in five of the six colon cancer samples. However, high levels of expression were observed in the matching normal adjacent tissue (NAT).

SEQ ID NO:26; Clone ID clnc00000551 (Sqcln054)

Semi quantitative PCR was done using the following primers:

Sqcln054forward:

5' AGCGGTCTCCTCCTCTTTCTAAA 3' (SEQ ID NO:66)

Sqcln054reverse:

5' CTGCCTCCTGATGCCTGAT 3' (SEQ ID NO:67)

Table 13 shows the absolute numbers which are relative levels of expression of sqcln054 in 12 normal samples from 12 different tissues. These RNA samples are individual samples or are commercially available pools, originated by pooling samples of a particular tissue from different individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 13: Sqcln054 in Normal Tissue Samples

Tissue	Normal
Breast	0
Colon	0
Endometrium	0
Kidney	0
Liver	0

5	Lung	0
	Ovary	0
	Prostate	0
	Small Intestine	0
	Stomach	0
	Testis	0
	Uterus	0

Relative levels of expression in Table 13 show that none of the normal tissues examined here exhibit expression of sqcln054.

Table 14 shows the absolute numbers which are relative levels of expression of sqcln054 in 12 cancer samples from 12 different tissues. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 14: Sqcln054 in Cancer Tissue Samples

20	Tissue	Cancer
	Bladder	0
	Breast	1
	Colon	0
	Kidney	0
25	Liver	0
	Lung	0
	Ovary	0
	Pancreas	1
	Prostate	0
30	Stomach	0
	Testes	0
	Uterus	0

Relative levels of expression in Table 14 show that sqcln054 is expressed in very low levels in breast, and pancreatic cancer. No expression was observed for colon cancer.

Table 15 shows the absolute numbers which are relative levels of expression of sqcln054 in 6 colon cancer matching samples. A matching pair is formed by mRNA from the cancer

sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in 5 duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 15: Sqcln054 in Colon Cancer Samples

10	Sample ID	Tissue	Cancer	NAT
	9609B019	Colon	1	0
	9709C074RA	Colon	0	0
	9705F002D	Colon	1	0
	9608B012	Colon	1	0
15	4004709A1	Colon	0	0
	9707C004GB	Colon	0	0

Relative levels of expression in Table 15 show that sqcln054 is expressed in low levels in three of the six colon cancer samples with no expression in the matching normal adjacent 20 tissue (NAT).

SEQ ID NO:41; Clone ID clnc00000614 (Sqcln077)

Semi quantitative PCR was done using the following primers:

Sqcln077forward:

25 5' TGTTCCTGAGCCAAAGATTGTAT 3' (SEQ ID NO:68)

Sqcln077reverse:

5' TGTGCCAGATACGAATGATGTTAC 3' (SEQ ID NO:69)

Table 16 shows the absolute numbers which are relative levels of expression of sqcln077 in 12 normal samples from 12 30 different tissues. These RNA samples are individual samples or are commercially available pools, originated by pooling samples of a particular tissue from different individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in 35 duplicate. Relative expression levels of 0, 1, 10, 100 and

1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 16: Sqcln077 in Normal Tissue Samples

5	Tissue	Normal
	Breast	0
	Colon	0
	Endometrium	0
	Kidney	0
10	Liver	0
	Lung	0
	Ovary	0
	Prostate	0
	Small Intestine	0
15	Stomach	0
	Testis	0
	Uterus	0

Relative levels of expression in Table 16 show no expression of sqcln077 in any of the normal tissues examined here.

20 Table 17 shows the absolute numbers which are relative levels of expression of sqcln077 in 12 cancer samples from 12 different tissues. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression
25 levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 17: Sqcln077 in Cancer Tissue Samples

30	Tissue	Cancer
	Bladder	0
	Breast	1
	Colon	1
	Kidney	0
	Liver	0
35	Lung	1
	Ovary	1
	Pancreas	0
	Prostate	0
	Stomach	0
40	Testes	0
	Uterus	1

Relative levels of expression in Table 17 show that sqcln077 is expressed in very low levels in colon, breast, lung, ovary and uterine carcinomas.

Table 18 shows the absolute numbers which are relative levels of expression of sqcln077 in 6 colon cancer matching samples. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 18: Sqcln077 in Colon Cancer Samples

Sample ID	Tissue	Cancer	NAT
9609B019	Colon	0	0
9709C074RA	Colon	0	0
9705F002D	Colon	0	0
9608B012	Colon	0	0
4004709A1	Colon	1	0
9707C004GB	Colon	1	0

Relative levels of expression in Table 18 show that sqcln077 is expressed in low levels in two of the six colon cancer samples with no expression in the matching normal adjacent tissue (NAT).

SEQ ID NO:50; Clone ID clnc00000078 (sqcln066)

Semi quantitative PCR was done using the following primers:

Sqcln066forward:

5' AGCAAAACAAGAGGCAGACAC 3' (SEQ ID NO:70)

Sqcln066reverse:

5' CAGGACACTCTGCCCTTGATA 3' (SEQ ID NO:71)

Table 19 shows the absolute numbers which are relative levels of expression of sqcln66 in 12 normal samples from 12 different tissues. These RNA samples are individual samples

or are commercially available pools, originated by pooling samples of a particular tissue from different individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in 5 duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 19: Sqcln66 in Normal Tissue Samples

10	Tissue	Normal
	Breast	0
	Colon	1
	Endometrium	1
	Kidney	1
15	Liver	0
	Lung	0
	Ovary	1
	Prostate	1
	Small Intestine	1
20	Stomach	1
	Testis	1
	Uterus	1

Relative levels of expression in Table 19 show that normal colon, endometrium, kidney, ovary, prostate, small intestine, 25 stomach, testis and uterus exhibit low levels of sqcln066 expression.

Table 20 shows the absolute numbers which are relative levels of expression of sqcln066 in 12 cancer samples from 12 different tissues. Using Polymerase Chain Reaction (PCR) 30 technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

35 Table 20: Sqcln66 in Cancer Tissue Samples

Tissue	Cancer
Bladder	0
Breast	1
Colon	0

5	Kidney	0
	Liver	0
	Lung	10
	Ovary	100
	Pancreas	1
	Prostate	10
	Stomach	10
	Testes	10
	Uterus	1

10 Relative levels of expression in Table 20 show that sqcln066 is expressed in very low levels in breast, pancreatic and uterine cancers. Moderate levels of expression are seen in lung, prostrate, testicular, ovary and stomach cancers.

Table 21 shows the absolute numbers which are relative
 15 levels of expression of sqcln066 in 6 colon cancer matching samples. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. Using Polymerase Chain Reaction (PCR) technology expression
 20 levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

25 **Table 21: Sqcln66 in Colon Cancer Samples**

30	Sample ID	Tissue	Cancer	NAT
	9609B019	Colon	0	0
	9709C074RA	Colon	1	0
	9705F002D	Colon	0	0
	9608B012	Colon	0	0
	4004709A1	Colon	0	0
	9707C004GB	Colon	10	0

Relative levels of expression in Table 21 show that sqcln066 is expressed in low to moderate levels in two of the six colon
 35 cancer samples with no expression in the matching normal adjacent tissue (NAT).

SEQ ID NO:52; Clone ID clnc00000101 (sqcln070)

Semi quantitative PCR was done using the following primers:

Sqcln070forward:

5' CAGCCTGCTACCACAGAACTCT 3' (SEQ ID NO:72)

5 Sqcln070reverse:

5' CCGTGCCTGATCTTTTGATAAGT 3' (SEQ ID NO:73)

Table 22 shows the absolute numbers which are relative levels of expression of sqcln070 in 12 normal samples from 12 different tissues. These RNA samples are individual samples or are commercially available pools, originated by pooling samples of a particular tissue from different individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 22: Sqcln070 in Normal Tissue Samples

	Tissue	Normal
20	Breast	0
	Colon	1
	Endometrium	0
	Kidney	1
	Liver	0
25	Lung	1000
	Ovary	0
	Prostate	100
	Small	1
	Stomach	0
30	Testis	10
	Uterus	0

Relative levels of expression in Table 22 show that normal testis prostate and lung exhibit moderate to high levels of expression of sqcln070. However, low levels of expression are noticed for normal tissues from colon, kidney and small intestine.

Table 23 shows the absolute numbers which are relative levels of expression of sqcln070 in 12 cancer samples from 12

different tissues. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 23: Sqcln070 in Cancer Tissue Samples

Tissue	Cancer
Bladder	0
Breast	1000
Colon	0
Kidney	1000
Liver	1
Lung	1
Ovary	10
Pancreas	1
Prostate	10
Stomach	1
Testes	1000
Uterus	1

Relative levels of expression in Table 23 show that sqcln070 is expressed in very low levels in liver, lung, pancreatic, stomach and uterine cancers. High expression levels for sqcln070 are seen in breast, kidney and testicular cancers, while moderate levels of expression is noticed in ovary and prostate cancers.

Table 24 shows the absolute numbers which are relative levels of expression of sqcln070 in 6 colon cancer matching samples. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 24: Sqcln070 in Colon Cancer Samples

Sample ID	Tissue	Cancer	NAT
9609B019	Colon	0	1000
9709C074RA	Colon	0	0
9705F002D	Colon	0	0
9608B012	Colon	0	0
4004709A1	Colon	1000	0
9707C004GB	Colon	0	0

Relative levels of expression in Table 24 show that sqcln070 is expressed in high levels in one of the six colon cancer samples and in another NAT of a different colon cancer paired sample.

SEQ ID NO:53; Clone ID clnc00000108 (Sqcln071)

Semi quantitative PCR was done using the following primers:

Sqcln071forward:

5' CTGTGCCCATTCTTACCAGTAA 3' (SEQ ID NO:74)

Sqcln071reverse:

5' GAGCCTTACTTTTTTTCTGCTGTA 3' (SEQ ID NO:75)

Table 25 shows the absolute numbers which are relative levels of expression of sqcln071 in 12 normal samples from 12 different tissues. These RNA samples are individual samples or are commercially available pools, originated by pooling samples of a particular tissue from different individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 25: Sqcln071 in Normal Tissue Samples

	Tissue	Normal
5	Breast	0
	Colon	0
	Endometrium	0
	Kidney	0
	Liver	0
10	Lung	0
	Ovary	0
	Prostate	0
	Small Intestine	0
	Stomach	0
	Testis	0
	Uterus	0

15 Relative levels of expression in Table 25 show that sqcln071 is not expressed in any of the normal tissue samples analyzed here.

Table 26 shows the absolute numbers which are relative levels of expression of sqcln071 in 12 cancer samples from 12 different tissues. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample 25 indicates the highest relative expression value.

Table 26: Sqcln071 in Cancer Tissue Samples

	Tissue	Cancer
30	Bladder	0
	Breast	0
	Colon	0
	Kidney	0
	Liver	0
35	Lung	0
	Ovary	0
	Pancreas	0
	Prostate	0
	Stomach	1
	Testes	1
	Uterus	0

Relative levels of expression in Table 26 show that sqcln071 is expressed in very low levels in stomach and testicular cancers.

Table 27 shows the absolute numbers which are relative levels of expression of sqcln071 in 6 colon cancer matching samples. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value

Table 27: Sqcln071 in Colon Cancer Samples

Sample ID	Tissue	Cancer	NAT
9609B019	Colon	1	0
9709C074RA	Colon	0	0
9705F002D	Colon	1	0
9608B012	Colon	0	0
4004709A1	Colon	0	0
9707C004GB	Colon	1	0

Relative levels of expression in Table 27 show that sqcln071 is expressed in low levels in three of the six colon cancer samples with no expression in the matching normal adjacent tissue (NAT).